

# Effects of Organizationally Endorsed Coaching on Performance and Validity of Situational Judgment Tests

Melissa S. Stemig\*, Paul R. Sackett\*\* and Filip Lievens\*\*\*

\*Department of Psychology, University of Minnesota, N498 Elliott Hall, 75 East River Road, Minneapolis, MN 55455, USA. stem0059@umn.edu

\*\*Department of Psychology, University of Minnesota, N475 Elliott Hall, 75 East River Road, Minneapolis, MN 55455, USA

\*\*\*Department of Personnel Management and Work and Organizational Psychology, Ghent University, Dunantlaan 2, 9000 Ghent, Belgium

**There is growing interest in organizationally provided or organizationally endorsed coaching. However, little is known about the effects of such coaching on test scores in operational settings. This study reports on an examination of such a program in the context of the use of a situational judgment test (SJT) for medical school admissions. We examine the effects of multiple types of coaching methods on SJT scores and on their construct-related and predictive validities. Results suggest that (1) commercial coaching techniques may not be as effective as previously thought, whereas organizationally provided methods may be more effective, and that (2) the criterion-related validity of the SJT scores is not degraded by the availability of coaching. Generally, this study illustrates that concerns about potential unfairness of coaching can be countered by making effective coaching available to all examinees, in the form of organizationally endorsed coaching.**

## 1. Introduction

Tests used in high-stakes situations such as employment and college admissions are frequently the subject of commercially available coaching programs that are intended to improve examinee scores in an effort to increase the likelihood of acceptance. Such programs are commonly viewed as a source of unfairness, with the assertion made that those with more financial resources are more able to access these coaching programs and thus obtain a competitive advantage. One response to this is the suggestion that coaching be made broadly and freely available to all, thus addressing the disparity in access to coaching.

A highly publicized recent example is the March 2014 announcement by The College Board of a multipronged plan for revision of the SAT college entrance examination. The Board acknowledged that 'the rise of costly test preparation has introduced a level of inequality among students', and thus in addition to changes to the content of the test, the Board announced a partnership with the Khan Academy to provide online access to

free test preparation and practice materials (College Board, (2014) (<https://www.collegeboard.org/delivering-opportunity/sat/practice>).

Thus, while there is growing interest in organizationally provided or organizationally endorsed coaching, there is a dearth of research on such coaching in operational settings. We report an examination of such a program in the context of the use of a situational judgment test (SJT) for medical school admissions in Belgium. The Belgian admissions system is quite different from that in other countries like the United States. A key feature is that admission is controlled by the government, rather than by universities. A government-run examination is administered annually, with multiple components (including general cognitive ability, science knowledge, and an SJT; data available for the present study are limited to the SJT; see Lievens, Buyse, & Sackett, 2005a, for a more detailed description of the other components). The government controls the number of students who pass the examination and are offered admission. Admitted students can attend any medical school in the country; schools do not accept or reject students. Hence,

medical schools' interactions with students focus on recruiting, rather than selection, as schools must attract the number of students they view as needed to justify and sustain their medical training program (after these students have passed the government-run admission exam).

It is in this context that a variety of types of coaching programs have emerged. First, commercial coaching programs have emerged. In the present study, we differentiate between two types of commercial coaching, namely programs administered online and the use of individual tutors. As the terms 'commercial' coaching implies, students have to pay to attend these coaching programs. Our particular interest, though, is in two other types of formal test preparation that are made freely available to all students by test developers and medical schools. The first is a booklet about the SJT produced by the developers. This booklet explains the purpose of the SJT and presents sample items and practice items (i.e., about 10 items, each with the correct responses). The second consists of information sessions organized by individual medical schools. As schools wish to attract students who pass the government-run examination, they have initiated a process of offering test information sessions on-site at the school. Similar to the booklet, these sessions also explain the SJT process and offer sample/practice items.

In the present study, students completing the medical school entrance examination self-reported whether or not they participated in any of the four types of formal coaching outlined earlier, as well as their participation in more informal test preparation activities, such as informal tutoring (by friends and family) and participation in online discussions/forums about the SJT. Among students taking the examination, we examine SJT differences between those participating in and not participating in each type of coaching; among the subset of test takers who pass the examination and attend medical school, we examine the relationship between coaching and academic performance in interpersonally oriented courses, and the effects of coaching on the criterion-related validity of SJT scores.

## 2. SJTs: overview of prior coaching research

SJTs are measures in which an applicant is presented with a set of hypothetical situations and then chooses among multiple presented courses of action. While the SJT is a measurement method that can be used to assess a variety of constructs, the SJT examined here focuses on two broad groups of interpersonal skills (communication and relationship building), which is a common use of SJTs (Christian, Edwards, & Bradley, 2010; Lievens et al., 2005a). Meta-analytic results (McDaniel, Morgeson,

Finnegan, Campion, & Braverman, 2001) indicate that SJTs perform well as predictors of job-related performance (0.26 uncorrected; 0.34 corrected) while simultaneously providing incremental validity over cognitive ability, personality, and the combination of the two (McDaniel, Psotka, Legree, Yost, & Weekley, 2007). Lievens (2013) recently demonstrated that video-based SJTs can function particularly well (better than cognitive ability) in predicting future interpersonal behaviors in clinical practice.

As the use of SJTs in large-scale testing programs continues to expand, it is essential to assess the degree to which coaching techniques can improve test scores. If certain coaching techniques are more effective than others, especially commercial programs, fairness of testing becomes an issue. Alternate scoring methods for SJTs have been developed to control for the effects of one specific coaching strategy in Black and White populations (i.e., avoiding extreme responses in rating the effectiveness of courses of action; McDaniel, Psotka, Legree, Yost, & Weekly, 2011). However, thus far, we have no evidence regarding the susceptibility of SJTs to a broad range of coaching activities such as the ones examined in this study.

In an experimental study, Cullen, Sackett, and Lievens (2006) developed test-taking strategies for two SJTs (College Student Questionnaire, CSQ (Sternberg and the Rainbow Project Collaborators, 2006), and the Situational Judgment Inventory, (SJI; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004) and provided coaching designed to mimic that provided by commercial coaching programs to undergraduate participants. Their results suggested that the CSQ was susceptible to two forms of coaching effects – instruction on what a 'good' answer is and advice to avoid extreme response patterns. Students in the coaching condition scored better than those that did not receive coaching (by 0.24 SDs). The SJI showed no coaching effect; participants in the coaching condition performed worse (0.22 SDs) than those in the control group, suggesting that SJTs coachability may vary across measures. As this was an experimental study, the impact of coaching on criterion-related validity was not examined.

More recently, Lievens, Buyse, Sackett, and Connelly (2012) assessed the effects of commercial coaching on SJT scores as part of a selection system for admission to medical school in Belgium. In order to estimate the treatment effects, the researchers compared the pre- and post-coaching SJT scores of individuals who had failed the overall exams once, sought commercial coaching, and then retaken the exams at a later date. With this methodology, results suggested that attending a commercial coaching program can improve SJT scores greatly ( $d = 0.59$ ) between the first and second examinations. This study also did not present evidence regarding the effects of coaching on subsequent outcomes such as

improved performance on the construct of interest later on or on criterion-related validity of SJT scores.

The present study builds upon the scarce body of research on SJT coaching in three ways. First, we make a distinction between different coaching strategies, as many methods are available to students besides commercial coaching and warrant consideration. Second, we examine whether the coaching received later transfers to enhanced interpersonal skills performance in medical school (as measured subsequent to selection). Finally, we determine whether the coaching effect is real or artificial. That is, does coaching lower the validity of the SJT scores?

### 3. Coaching effects on validity

Messick (1981) describes three ways in which coaching can affect test scores (see also Lievens, Buyse, & Sackett, 2005b). The first is to increase an examinee's score without affecting the individual's standing on the underlying construct of the measure. This first effect negatively impacts both the construct-related and criterion-related validity of test scores as the scores for coached individuals will be artificially inflated, indicating a higher-than-true score on the variable of interest, and the accuracy of predictions made from these scores could decrease. The second effect is to increase scores for examinees who, without coaching, would score lower than their true score because of anxiety of unfamiliarity with the exam itself. This second effect positively affects both the construct-related and criterion-related validity of test scores. The third effect is to genuinely improve the examinee's standing on the underlying construct itself. This would have no effect on either the construct or criterion-related validity of the test scores.

Underlying theory can clarify the various effects coaching may have on SJT scores and their validity. Motowidlo and Beier (2010) posited a theory which connects personality, ability, experience, specific job knowledge, and implicit trait policies (ITPs; Motowidlo, Hooper, & Jackson, 2006a, 2006b) to job performance. The theory suggests that SJTs measure the procedural knowledge that consists of specific job knowledge and ITPs. ITPs are the implicit beliefs held by an individual about the relationships between the expression of certain personality traits and effectiveness in job situations. For example, an individual may believe that extraverted action is generally more effective than introverted action, but they learn that in certain job situations it is more effective to display introversion and that in others extraversion is irrelevant for effectiveness.

If personality and ability are assumed to be unchangeable in the timeframe of a brief coaching intervention, then coaching can potentially affect only one input in

this model: ITPs. According to the model, a result of improving ITPs would be an increase in job performance – this corresponds to Messick's (1981) third effect and would have no effect on either type of validity. However, if coaching does not improve subsequent job performance, this would suggest that coaching influences a variable outside of procedural knowledge and therefore would decrease the construct-related validity of test scores.

## 4. Method

Participants were 1,085 individuals who took a series of exams for entry into medical schools in Belgium for the first time in July of 2008, and who responded to a survey asking them about their test preparation activities. The survey was presented as for research purposes only, and anonymity was assured.

### 4.1. Measures

#### 4.1.1. Survey

Respondents indicated whether they had participated in 11 types of coaching procedures (specifically for the SJT portion of the exam). To collect a representative set of coaching activities, these 11 coaching procedures were determined on the basis of prior research (e.g., Becker, 1990; Powers & Rock, 1999), interviews with students, and web searches. Seven of these coaching procedures were examined in the present study because the other coaching types had very low rates of student participation. These seven coaching procedures were grouped into three categories in light of the distinction made earlier:

- Free officially endorsed coaching methods:
  - Attending information sessions at university (test-taking skills and completing sample test items)
  - Completing the sample test items in the official brochure
- Commercial, unsanctioned coaching methods:
  - Online commercial coaching methods
  - In-person commercial tutoring
- Other coaching methods
  - Reading books about how to take the tests
  - Informal tutoring by friends or family members
  - Utilizing free online resources (e.g., forum discussions, websites)

In addition to the coaching procedures, several demographic characteristics were collected. The examinees reported parental socio-economic status (SES; mother's and father's education) as well as their quartile rank in high school. They also completed a self-report item asking about the degree to which attending medical school would impose financial difficulty. All variables were self-reported by the student.

#### 4.1.2. SJT score

The SJT was designed to assess interpersonal skills (communicating information and relationship building) in a physician–patient interaction. Examinees responded to 30 knowledge-based questions (“What is the best response?”). Responses were scored using a key developed by medical expert agreement (see Lievens et al., 2005b, for complete descriptions). Scores were available for all participants in the dataset.

#### 4.1.3. Interpersonal skills grade point average (GPA)

While at medical university, students take several interpersonally oriented courses – courses that are focused on instructing and improving interpersonal and communication skills. Exams and exercises of these courses are typically interactional in nature. Interpersonal GPA was calculated by retrieving students’ scores in the interpersonal course from the archives. The current dataset contained all individuals who responded to the questionnaire during the July examination – including those who did not pass and therefore did not attend medical school (the examination has roughly a 30% pass rate). Interpersonal GPA could be retrieved only from medical schools that had interpersonal courses as formal requirements in their curriculum, resulting in a sample size of 154 for analyses of SJT validity. In light of possible differences in course content and grading, scores were standardized within medical school.

## 5. Results

### 5.1. Participation in coaching

Table 1 presents means, standard deviations, and intercorrelations among study variables. Large proportions of the examinees took advantage of the freely available coaching methods (e.g., 69% completed the exercises in the official brochure, and 41% attended information sessions at a university), whereas fewer utilized commercial coaching techniques (e.g., 8% used private tutoring and 18% used paid online coaching). There is no evidence of a substitution effect for coaching procedures (i.e., using one coaching method does not mean others are not used), as all correlations between coaching variables are positive. Most correlations are modest in magnitude. One outlier is the correlation of .63 between paid online coaching and using free online resources. This is probably due to the fact that people who paid for online coaching also surfed the Internet for freely available online coaching resources. It is also possible that the online coaching programs provided them with links to other free online resources.

Table 1 also provides information as to whether socioeconomic status and prior academic performance were related to participation in various coaching activities. Our initial focus was on commercial coaching (paid

websites and individual tutoring). Parental education was positively related to the use of tutoring, and the degree to which medical school would be a financial burden was negatively related to tutoring. These findings support the general fear that commercial coaching is a resource more available to those with greater financial resources. The one other strong pattern of relationships was with attending information sessions at medical schools. Again, parental education was positively related, and viewing medical school as a financial burden was negatively related. In addition, stronger high school performance was related to attending these sessions. Thus even though these information sessions were freely available to all, students with stronger academic records and those with higher parental SES were more likely to take advantage of this opportunity.

### 5.2. Effectiveness of coaching

As an initial examination of the effects of the coaching procedures on subsequent SJT scores, Cohen’s  $d$  was computed between those that received treatment and those that did not. Table 2 presents the results. One initial surprising result was that the use of paid tutoring had a negative effect on SJT scores ( $d = -0.19$ ). This descriptive analysis, however, does not control for the possibility of selection bias. It is feasible that those who sought coaching would have in fact scored even lower if they had not been coached, and that paid coaching is indeed effective. We attempted to address this selection effect by controlling for key covariates: previous academic performance and parental socioeconomic status (a combination of maternal and paternal educational attainment). However, controlling for these variables had a negligible effect. Table 2 does show that other types of coaching were linked to differences in SJT scores. Attending information sessions at university ( $d = 0.51$ ) and completing the exercises in the official test brochure ( $d = 0.39$ ) produced significant effects (significance is based on the correlation between participation in these types of coaching and SJT scores as reported in Table 1). Again, there was a concern for selection effects and therefore we controlled for parental SES and high school GPA. After these controls are in place, the effect of attending information sessions was reduced (to  $d = 0.46$ ), but still significant. Similarly, the effect of using the exercises in the official brochure was reduced (to  $d = 0.37$ ), but not to a meaningful degree.

### 5.3. Effects on validity

#### 5.3.1. Construct-related validity of SJT scores

The previous evidence suggests that coaching can improve SJT scores, but it is not certain whether coaching improves test scores alone or if they also improve subsequent interpersonal skills GPA as assessed in

Table 1 Means, SDs, and intercorrelations among study variables

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12
Predictor														
1. SJT	15.37	3.52	–											
Free officially endorsed coaching methods														
2. Info sessions	0.41	0.49	0.25**	–										
3. Exercises in brochure	0.69	0.47	0.18**	0.19**	–									
Commercial, unsanctioned, coaching														
4. Paid websites	0.18	0.38	0.06*	0.08**	0.22**	–								
5. Paid tutor	0.08	0.26	–0.05	0.08**	0.04	0.13**	–							
Other coaching methods														
6. Free websites	0.36	0.48	0.04	0.12**	0.34**	0.63**	0.09**	–						
7. Read books	0.20	0.40	–0.00	0.13**	0.22**	0.22**	0.12**	0.27**	–					
8. Informal tutoring	0.40	0.49	0.06	0.13**	0.36**	0.14**	0.12**	0.31**	0.31**	–				
Control covariates														
9. Education father <sup>a</sup>	4.82	0.71	0.11	0.06*	–0.02	0.02	0.08*	–0.03	–0.02	0.03	–			
10. Education mother <sup>a</sup>	4.53	1.41	0.14**	0.11**	0.05	–0.00	0.08**	–0.03	0.02	0.05	0.57**	–		
11. Financial burden <sup>b</sup>	1.98	1.00	–0.03	–0.09**	–0.01	–0.06	–0.10**	–0.07*	–0.01	–0.06	–0.31**	–0.26**	–	
12. High school rank <sup>c</sup>	1.55	1.33	–0.16**	–0.06*	–0.06*	0.02	–0.02	–0.03	0.02	0.02	–0.02	–0.09**	0.01	–
Criterion														
13. Interpersonal GPA <sup>d</sup>	0.12	1.00	0.17*	0.13	0.15	0.06	0.14	0.00	0.09	–0.01	0.07	0.10	0.11	–0.22**

\*Significant at the 0.05 level (two-tailed). \*\*Significant at the 0.01 level (two-tailed).

<sup>a</sup>Measured on a 1–6 scale: 1 = only primary education, 2 = first 3 years of secondary education, 3 = completed secondary education, 4 = college, but not university (short-type), 5 = college, but not university (long-type), 6 = university.<sup>b</sup>Measured on a 1–5 scale: 1 = no financial problem, 2 = relatively easy, 3 = neutral, 4 = difficult, 5 = huge financial burden.<sup>c</sup>Measured on a 1–4 scale: where 1 is the top quartile.<sup>d</sup>N = 154.

GPA = grade point average; SD = standard deviation.



Table 2. Coaching participation rates and differences in SJT score for coached versus uncoached individuals

	N (yes)	N (no)	M (yes)	M (no)	SD (yes)	SD (no)	d (yes–no)
Free officially endorsed coaching methods							
Info sessions	428	613	16.41	14.65	3.27	3.52	0.51
Exercises in brochure	719	327	15.79	14.44	3.38	3.67	0.39
Commercial, unsanctioned, coaching							
Paid websites	187	851	15.82	15.26	3.28	3.58	0.16
Paid tutor	78	963	14.72	15.41	3.63	3.51	–0.19
Other coaching methods							
Free websites	376	671	15.53	15.28	3.40	3.59	0.07
Read books	206	836	15.33	15.37	3.51	3.53	–0.01
Informal tutoring	415	629	15.62	15.21	3.47	3.54	0.12

SD = standard deviation; SJT = situational judgment test.

Table 3. Regression of interpersonal GPA on coaching and control variables

	Unstandardized coefficients		Standardized coefficients	
	B	SE	$\beta$	t
(Intercept)	–2.66	0.81		–3.27**
Free officially endorsed coaching methods				
Info sessions	0.34	0.18	0.16	1.89
Exercises in brochure	0.42	0.19	0.20	2.24*
Commercial, unsanctioned, coaching				
Paid websites	0.12	0.27	0.05	0.44
Paid tutor	0.74	0.38	0.16	1.95
Other coaching methods				
Free websites	–0.18	0.25	–0.09	–0.74
Read books	0.32	0.25	0.11	1.31
Informal tutoring	–0.06	0.18	–0.03	–0.33
Control covariates				
Education father	0.06	0.07	0.07	0.76
Education mother	0.05	0.08	0.06	0.65
Financial burden	0.20	0.09	0.20	2.31*
High school rank	–0.27	0.10	–0.24	–2.89**

\*Significant at the 0.05 level. \*\*Significant at 0.01 level. GPA = grade point average; SE = standard error.

interpersonal courses during medical school. That is, does coaching produce an artificial increase in scores on an interpersonally oriented SJT, or does it also lead to real improvement in the construct of interest (interpersonal skills)? To answer this question, we regressed interpersonal skills GPA scores on the various coaching activities. The only significant effect present for the coaching variables was for completing the sample test items in the official brochure (see Table 3).

### 5.3.2. Predictive validity of SJT scores

The validity of the SJT in predicting GPA in interpersonal skills courses (0.169) was slightly reduced (to 0.153) in a model which controlled for the SJT coaching activities. When the analysis was limited to the two coaching activities that are most related to SJT scores (attending information session at the universities and completing the exercises in the official brochure), thus consuming fewer degrees of freedom, the SJT–Interpersonal Skills GPA relationship reduction is also attenuated (0.160).

We note that the predictive validity of the SJT in the sample of students for whom we have criterion data is lower than would be expected in the full applicant pool due to range restriction. The SJT standard deviation in the applicant pool is 3.52; the standard deviation in the validity study sample is 2.88. While the current dataset does not contain data on all variables used in the medical student selection process (which is needed for calculating indirect range restriction corrections), applying the standard correction for direct range restriction raises the SJT validity value from .169 to .203.

## 6. Discussion

Although large organizations (e.g., The College Board) are nowadays investing in organizationally endorsed coaching, little is known about its effectiveness. This study is the first to focus on organizationally endorsed coaching and compares its effectiveness with other coaching tactics in terms of affecting SJT scores and validities. Our results have a positive message: the most

effective coaching procedures were both freely available and officially sanctioned, and involved practice with items similar to those on the actual SJT. In contrast, informal methods (e.g., tutoring from friends) are less effective. It appears that coaching is most effective when students get a chance to see similar items to those on the actual SJT, and to understand what makes a good answer. These methods are similar to the kinds of things done in commercial programs in the United States, but fairness is less of an issue in the present case as they are freely available to all students. Therefore, these results do indicate that there is a possibility of improving SJT scores via coaching methods.

We note, however, that simply making effective coaching freely available may not be sufficient. We found that students with better academic records and with greater parental education were more likely to take advantage of the information sessions at medical schools. We believe that the correlation between SES and free coaching merits further investigation into the mechanisms underlying this relationship. For example, higher SES parents may value college attendance more highly, and may thus exert more effort to encourage their children to do all they can to improve their chances of admission. We suggest that extra effort will be needed to make less advantaged students aware of the value of these coaching opportunities.

Regarding commercial coaching effects, the results from this study are in contrast to those found in Lievens et al. (2012); they found that there was a substantial effect size present for commercial coaching procedures in a sample of students who had failed the exam once ( $d = 0.59$ ). Our study differentiated between paid online coaching and individual tutoring, while Lievens et al. combined all forms of commercial coaching. We found an effect of online coaching ( $d = 0.16$ ), but a surprising negative effect for individual tutoring ( $d = -0.19$ ). Tutoring is not a regulated or credentialed industry; it appears that there are individuals offering services who are not equipped to do so. One possible explanation for the larger effect in Lievens et al. is that failure motivates students to attend more carefully to a coaching program, producing different effects than when coaching is prior to any test. The basis for the difference between the current finding and prior findings (Lievens et al., 2012) remains to be resolved.

Additionally, a key finding of this study is that coaching received does not appear to transfer into interpersonal skill performance. That is, most coaching methods are effective at raising the score on the SJT, but do not necessarily improve interpersonal skills. The exception to this finding is completion of the sample items in the test brochure – this we interpret not as an impact on the construct itself (i.e., completing the sample items in the brochure increases interpersonal skills), but rather

posit that it may be due to a conscientiousness effect: those who do not read the brochure are also likely to be unreliable in other domains, leading to a lower performance. The difference, then, may be attributed to a sample selection effect rather than real coaching effectiveness. This suggestion, however, warrants further investigation.

Importantly, our results also indicate the validity of SJTs as predictors of interpersonal skills is not diminished in the presence of coaching procedures. This finding is particularly useful to practitioners using the SJT as a predictive measure today.

Our research does offer support for the concerns about inequality in access to coaching, as we find modest, but statistically significant, relationships between indicators of socioeconomic status and participation in commercial coaching. We also find relationships between SES indicators and SJT scores. However, controlling for SES does not produce appreciable change in the predictive power of the SJT, a finding that parallels research with traditional cognitively focused admissions tests (Sackett, Kuncel, Arneson, Cooper, & Waters, 2009). The effects of making access to coaching available to all test takers, and investigating mechanisms to increase the rate with which students take advantage of such coaching opportunities are issues meriting further research.

The current research takes place in one specific context – medical school admissions in Belgium – and examination of these issues in other contexts and cultures is needed. As the present study took place in an admission context, implications for coaching in the context of personnel selection merit discussion. It is, of course, the case that research in such contexts would be useful; we offer the present study as an initial foray into the study of organizationally sanctioned coaching. We note that the applicability of the issue of organizationally sanctioned coaching to employment testing will vary by the type of setting. We posit that coaching issues will be most relevant in settings where there are concerns that external agents are providing coaching, with the possibility of some examinees have more ready access to coaching than others. Such settings generally involve jobs that are viewed as highly desirable, such that candidates are motivated to invest time and effort in test preparation and thus generate high numbers of applicants. Examples might include large-scale public-sector testing programs, such as police or fire testing in large cities, and highly visible federal job settings, such as the US Postal Service. In other settings, testing programs may be less visible, testing volumes may be too small to attract commercial coaching, or the job in question may not be sufficiently attractive that candidates are willing to invest in preparation activities.

One limitation to this research is that while we controlled for several key variables in determining

likelihood to seek coaching (SES, high school rank), there may be other variables present that were not controlled for. Coaching is generally a difficult phenomenon to study, as random assignment to conditions is not feasible in high-stakes settings. Candidates motivated to seek coaching will do so. Thus, the use of control variables, as in the present study, or propensity scores, as in the case of Lievens et al. (2012) and Connelly, Sackett, and Waters (2013), appear to be the best available strategies for addressing the selection bias issue.

Another limitation is that our survey might not have captured all relevant coaching activities, even though the coaching activities included were determined on the basis of multiple sources (i.e., prior research, web searches, and interviews with students). Relatedly, our study did not address differential intensity of student investment in a particular type of coaching. The student survey only asked about the types of coaching in which a student engaged, not the number of times or hours incurred in a given method. Thus, even though a certain student may have attended two information sessions while another only attended one, they appear the same in our analysis. The present study offers useful insights for operational testing programs. It illustrates countering concerns about potential unfairness of coaching by making effective coaching, in the form of school-sponsored coaching (e.g., information sessions and booklets containing exercises and sample items), available to all examinees. It also illustrates that the criterion-related validity of the SJT, modest though it is, is not materially degraded by the availability of coaching.

## References

- Becker, B. J. (1990). Coaching for the Scholastic Aptitude Test: Further synthesis and appraisal. *Review of Educational Research*, 60, 373–417.
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Construct assessed and a meta-analysis of their criterion related validities. *Personnel Psychology*, 63, 83–117. doi: 10.1111/j.1744-6570.2009.01163.x.
- College Board. (2014). The College Board announces bold plans to expand access to opportunity; redesign of the SAT. CollegeBoard News & Press. <https://www.collegeboard.org/releases/2014/expand-opportunity-redesign-sat>
- Connelly, B. S., Sackett, P. R., & Waters, S. D. (2013). Balancing treatment and control groups in quasi-experiments: An introduction to propensity scoring. *Personnel Psychology*, 66, 407–442.
- Cullen, M. J., Sackett, P. R., & Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment*, 14, 142–155.
- Lievens, F. (2013). Adjusting medical school admission: Assessing interpersonal skills using situational judgment tests. *Medical Education*, 47, 182–189.
- Lievens, F., Buyse, T., & Sackett, P. R. (2005a). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, 90, 442–452.
- Lievens, F., Buyse, T., & Sackett, P. R. (2005b). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology*, 58, 981–1007.
- Lievens, F., Buyse, T., Sackett, P. R., & Connelly, B. S. (2012). The effects of coaching on situational judgment tests in high-stakes selection. *International Journal of Selection and Assessment*, 20, 272–282.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Predicting job performance using situational judgment tests: A clarification of the literature. *Journal of Applied Psychology*, 86, 730–740.
- McDaniel, M. A., Psotka, J., Legree, P. J., Yost, A. P., & Weekley, J. A. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60, 63–91.
- McDaniel, M. A., Psotka, J., Legree, P. J., Yost, A. P., & Weekly, J. A. (2011). Toward an understanding of situational judgment item validity and group differences. *Journal of Applied Psychology*, 96, 327–336.
- Messick, S. (1981). The controversy over coaching: Issues of effectiveness and equity. In B. F. Green (Ed.), *New directions for testing and measurement: Issues in testing – coaching, disclosure, and ethnic bias*, no 11 (pp. 21–53). San Francisco, CA: Jossey-Bass.
- Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology*, 95, 321–333. doi: 10.1037/a0017975.
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006a). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, 91, 749–761. doi: 10.1037/0021-9010.91.4.749.
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006b). A theoretical basis for situational judgment tests. In J. A. Weekly, & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 57–81). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, 89, 187–207.
- Powers, D. E., & Rock, D. A. (1999). Effects of coaching on SAT I: Reasoning test scores. *Journal of Educational Measurement*, 36, 93–118.
- Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin*, 135, 1–22.
- Sternberg, R. J., & the Rainbow Project Collaborators. (2006). Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence*, 34, 321–350.